

Dealing with Frame Cancellation for Stereoscopic Displays in 3D User Interfaces

Jérémy Lacoche*

IRT b<>com

Thierry Duval ‡

Télécom Bretagne

IRT b<>com

Morgan Le Chénéchal*

IRT b<>com

Valérie Gouranton¶

INSA de Rennes

IRT b<>com

Sébastien Chalmé†

IRT b<>com

Eric Maisel §

ENIB

IRT b<>com

Jérôme Royan†

IRT b<>com

Bruno Arnaldi ¶

INSA de Rennes

IRT b<>com

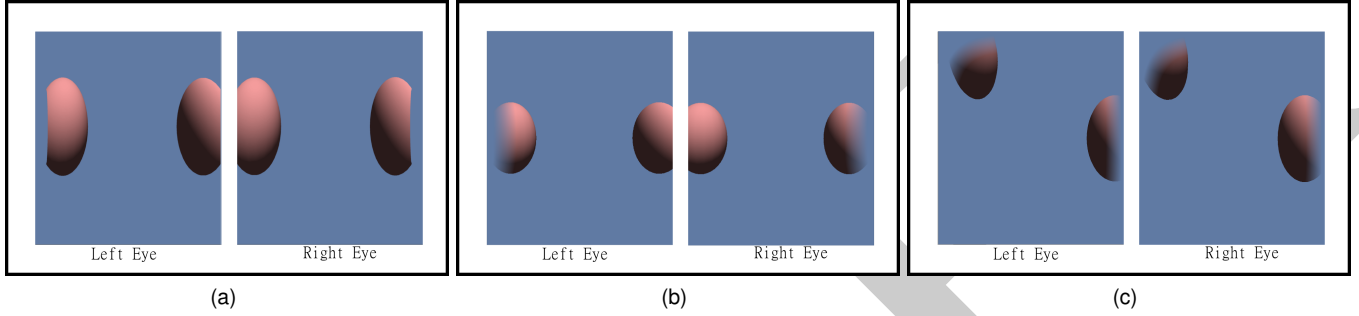


Figure 1: (a) SCVC approach illustration proposed by Ardouin et al. [1]. The clipping of the two spheres in negative parallax avoids inconsistent left and right images. (b) Progressive SCVC applied to the virtual objects close the screen edges. (c) Virtual alpha blended window effect applied to the virtual objects close the screen edges.

ABSTRACT

This paper aims at reducing the ocular discomfort created by stereoscopy due to the effect called “frame cancellation”, for movies and interactive applications. This effect appears when a virtual object in negative parallax (front of the screen) is clipped by the screen edges; stereopsis cue lets observers perceive the object popping-out from the screen while occlusion cue provides observers with an opposite signal. Such a situation is not possible in the real world. This explains some visual discomfort for observers and leads to a poor depth perception of the scene. This issue is directly linked to the physical limitations of the display size that may not cover the entire field of view of the observer.

To deal with these physical constraints we introduce two new methods in the context of interactive applications. The first method consists in two new rendering effects based on progressive transparency that aim to preserve the popping-out effect of the stereo. The second method focuses on adapting the interaction of the user, not allowing him to place virtual objects in an area subject to frame cancellation. Both methods have been evaluated and have shown a good efficiency in comparison to the state of the art approaches.

Index Terms: H.5.2 [Information interfaces and presentation]: User Interfaces—user centered design); I.3.3 [Computer Graphics]: Picture/Image Generation—display algorithms); I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Virtual Reality);

*{jeremy.lacoche, morgan.lechenechal}@b-com.com

†{sebastien.chalme, jerome.royan}@b-com.com

‡thierry.duval@telecom-bretagne.eu

§maisel@enib.fr

¶{bruno.arnaldi, valerie.gouranton}@irisa.fr

1 INTRODUCTION

Depth perception is the visual ability to perceive the world in three dimensions. The human visual system (HVS) perceives depth thanks to a variety of depth cues. These cues are separated into two categories. The first one contains monocular cues that are interpreted by each eye. For example, in this category we can find motion parallax, perspective, occlusion or texture gradient. The second category includes binocular cues that require both eyes to be perceived. Stereopsis and convergence are such cues.

Stereoscopic display is a sensorimotor contingency used for Virtual Reality (VR) applications in order to improve place and plausibility illusion [11]. Indeed, stereoscopy creates a three dimensional illusion by simulating stereopsis depth cue. However, in some cases, stereoscopy may also lead to visual discomfort. When virtual objects in negative parallax (front of the screen) are overlapped by one of the screen borders (horizontal and/or verticals), two depth cues conflict. The first one, stereopsis, allows depth perception with the ability of our visual system to merge the two different images acquired by the two eyes. In negative parallaxes, this cue tells our visual system that the object is in front of the screen. The second cue is occlusion; if a screen edge is partially overlapping an object, the edge seems to be closer. Moreover, because of the limited display size, one part of an object can be clipped for one camera and not for the other [10]. It is a particular case of binocular rivalry. Binocular rivalry occurs when the two eyes acquire two unrelated images. With horizontally aligned eyes, the parallax is only horizontal, and then only vertical edges are concerned by binocular rivalry. In case of a head tracking that considers head rotation around the 3 degrees of freedom, vertical parallax is also possible. Therefore, horizontal edges can also be concerned by binocular rivalry.

This phenomenon, first described by Valys [12] is called “frame cancellation” and is described in Figure 2. This conflict reduces the illusion of depth in the perceived 3D image [15] and creates an unpleasant effect sometimes called “eyestrain” for people watching the screen [8]. These consequences of frame cancellation have been observed in [1]. As frame cancellation sometimes induces

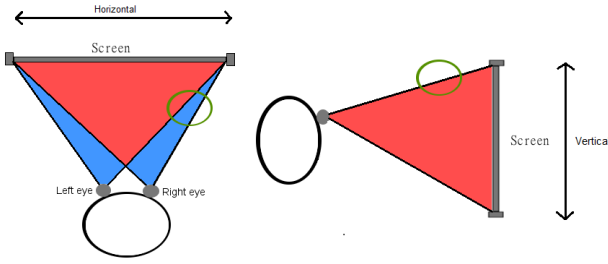


Figure 2: Top view and side view of camera volumes with horizontally aligned eyes. In red, the area without conflict. In blue, the area where objects will be visible only by one camera (binocular rivalry). The green circle represents a virtual object subject to frame cancellation. In case of non horizontally aligned eyes, the blue area could also be visible on the side view.

binocular rivalry, the problem is reinforced.

Frame cancellation is mainly caused by the physical limitations of most available visual displays: they do not fill the entire field of view of the observers. CAVE™ systems and head-mounted displays (HMD) are two display types that can avoid the problem by totally immersing the user in the virtual world thanks to a wide field of view. Nevertheless, when such a display is not available, we need to adapt the rendering and/or the interaction in order to reduce the frame cancellation effect.

In this paper we introduce two novel approaches to deal with frame cancellation. The first one aims at improving the state of the art techniques for the rendering of objects that are in the conflict zones. Our method is based on applying a progressive transparency for virtual objects that are in negative parallax and close to the screen borders. We propose two rendering effects smoother than the current ones such as the floating window [2] or the stereo compatible volume clipping (SCVC) [1]. Moreover, these effects offer a minimal reduction of the user's field of view and do not change the scene topology.

The second technique is especially designed for 3D user interfaces where virtual objects are manipulated by one user. Instead of adapting the rendering of objects in conflict zones, we propose to adapt the manipulation technique to avoid any virtual object to enter in these areas. The technique is based on a collision detection system with the two camera frustums that constrain the virtual objects during the interaction.

This paper is structured as follows: In Section 2 we introduce the existing methods that deal with frame cancellation and with other physical display constraints. In Sections 3 and 4 we detail our two approaches, the rendering effects and the interaction technique, respectively. The two approaches are evaluated in Section 5. Section 6 concludes the paper.

2 RELATED WORK

Adapting the rendering to deal with frame cancellation has been addressed for offline and real-time purposes. Indeed, 3D movies as well as interactive 3D applications can suffer from this problem.

One common approach consists of redefining the display window with virtual black bands on the screen borders when some objects are subject to the frame cancellation effect. The Floating Window introduced by Autodesk [2] for offline rendering and the Cadre Viewing approach proposed by Mulder et al. [10] for real-time contexts are solutions based on vertical black bands. The distance of the black bands is computed in order to maximize the field of view of the user while clipping the objects correctly to avoid the conflict between occlusion and stereopsis. The virtual objects are then perceived in positive parallax related to the virtual window, and no part of a virtual object in negative parallax is then visible by only one eye. The main drawback of this approach is field of view reduction.

Moreover, the occlusion issue with the horizontal screen borders is not addressed.

In the non-real time rendering field, some techniques consist of applying an image effect on the two final images in order to modify the disparity of the stereo pair globally or close to the screen borders. Such a method is introduced by Lipton et al. [8]. The approach consists of stretching the left and the right images close to the screen borders. The left part of the left image will be stretched as well as the right part of the right image. This effect gradually decreases the parallax values over the screen area close to the screen edges. Negative parallaxes are still ensured at the center of the screen in order to preserve the popping-out effect. The main drawback of this algorithm is the difficulty to apply it in real-time. Another drawback: The scale of the scene at the screen borders is altered. It may cause some disturbing effects to the observer and decrease the depth perception for these areas. In the same way, Lipton et al. [8] also describe a cross fading effect to locally modify the parallax values next to the vertical edges. Next to these sides, the corresponding pixels of left and right view are blended. As colors from the left and the right image can be blended, the final result may suffer from color inconsistency.

In a real-time context, the Cyclopean Scale introduced by Ware et al. [13] proposes to scale the virtual environment (VE) about a point between the observer's two eyes in order to always place it just behind the screen in terms of stereoscopic depth. This method avoids frame cancellation by always placing the VE in positive parallaxes or very small negative ones. Therefore, the Cyclopean Scale reduces the possibility of simulating popping-out effects offered by stereoscopic rendering. It is particularly limited in the case of co-located interactions.

One of the most relevant approaches has been proposed by Ardouin et al. [1]. The effect is called the Stereo Compatible Volume Clipping (SCVC) and consists of rendering only the part of the virtual objects that are in the conflict-free area: the Stereo Compatible Volume (SCV). This is done by clipping all the object's parts that intersect the blue area, as shown in Figure 2. One possible result is given in Figure 1a. This technique always ensures two consistent right and left images, does not reduce the field of view, and is fully compatible with head tracking. However all conflicts are not solved. Indeed, the technique mainly focuses on solving the binocular rivalry issue, and no part of an object in negative parallax is visible by only one eye, but may still be perceived as roughly clipped by a horizontal or by a vertical screen border, which disrupts the object form perception.

Most of the current methods focus on adapting the rendering to deal with frame cancellation. But, we can also find an approach that consists in adapting an interaction technique to reduce the frame cancellation effect. Wartell et al. [14] introduce a navigation technique for a large dataset based on scaling, panning, and rotating. During the navigation, the technique performs automatic translation of the scene perpendicular to the projection plane in order to always minimize frame cancellation. The translation is computed by detecting the collision between the scene and a plane slightly above the projection plane, and parallel to it, with the help of the depth buffer. Such a method can be compared to the Cyclopean Scale because it also avoids important negative parallaxes values. Therefore, it reduces the possibility of simulating popping-out effects.

3 ALPHA BLENDING TO AVOID FRAME CANCELLATION

The main drawbacks of the current methods are field of view reduction, scene scale alteration, or a too rough clipping of virtual objects. Moreover, no approach considers the horizontal edges of the screen in case of horizontally aligned eyes, while they may also occlude virtual objects in negative parallax.

That is why we introduce two approaches that aim at solving

these drawbacks by proposing a smoother technique based on alpha blending that will affect virtual objects in negative parallax. This alpha blending can be compared to the cross fading introduced in [8]. Instead of blending the left and right images, we propose to apply an alpha blending to create a transparency effect.

Both effects can be implemented in real-time and offline rendering engines. They are compared and evaluated in Section 5.

3.1 Progressive SCVC

The main drawback of the stereo compatible volume clipping is the rough clipping of virtual objects. This clipping may lead to visual discomfort of the observer, particularly when a virtual object appears behind a clipped one. In order to create a smoother effect, we propose to apply SCVC progressively with a clipping based on alpha blending.

To compute the alpha value applied to each pixel (x, y) , first we evaluate the distance $d(x, y)$ of the corresponding 3D point to the frame cancellation area. To do so, we compute the distance of this 3D point to the plane that is supposed to clip the object in SCVC. When we consider a user's head tracking around 3 degrees of freedom, eyes can be non-horizontally aligned. In that case, vertical parallax is possible, which means that the blue area visible in the horizontal view of Figure 2 will also be present in the vertical view of the same figure. That is why in that particular case, SCVC computes two planes per eye. This plane selection and computation is described more in [1]. In that case, we compute two distances, one for each plane, then the minimal distance is taken to compute the alpha value (0: totally transparent, 1: totally opaque). The alpha factor that linearly decreases when the 3D point gets closer to the plane is computed with respect to the equation 1.

$$\text{Alpha}(x, y) = 1 - \frac{d(x, y) - \text{distMinPlane}}{\text{distMaxPlane} - \text{distMinPlane}} \quad (1)$$

where distMinPlane is the distance where the transparency begins (maximal opacity) and distMaxPlane the distance where the transparency ends (minimal opacity). Both values are expressed in meters: distMinPlane can take a negative value in order to really clip the virtual object after the SCVC planes. The alpha value is then clamped between 0.0 and 1.0. A result is given in Figure 1b where we apply progressive SCVC on the two spheres subject to frame cancellation. For this image we have set distMinPlane to -0.02m and distMaxScreen to 0.05m and the sphere radius is 0.1m . Compared with SCVC in Figure 1a, we observe a smoother result that may lead to a better visual comfort in frame cancellation situations.

3.2 Virtual alpha blended window

As stereo compatible volume clipping [1], progressive SCVC considers the two horizontal edges only in case of head-tracking with non horizontally aligned eyes, and then mainly focuses on solving binocular rivalry created by frame cancellation. Even without head-tracking, the horizontal edges may occlude virtual objects in negative parallax and cause stereo cue conflicts. The virtual alpha blended window aims to take into account this issue by considering vertical edges of the screen as well as the horizontal ones in all cases. This approach consists of applying a progressive alpha blending over the screen edges to the virtual objects in negative parallax for the two cameras. Contrary to SCVC and progressive SCVC, with this effect, when an object is close to a screen edge, its rendering is adjusted in the views of both the left and right eye. Indeed, we suppose that modifying only one view can lead to an increased visual discomfort and difficulty to fuse the two images.

As for progressive SCVC, we propose applying a linear decreasing alpha blending, but this time, on the two axes of the screen. To apply the effect, we consider the screen as a space where all pixel coordinates take a value between $(-1, -1)$ (bottom left corner) and

$(1, 1)$ (top right corner). The transparency is computed in order to get a "squire" shape, a particular case of a superellipse. A square shape would have been coarse on the screen corners, and a circle shape would erase too much of the objects on these corners. For both cameras, for a pixel (x, y) , we compute its alpha factor only if the corresponding position in the 3D world is in front of the screen (negative parallax). Concerning the alpha factor computation, we first compute its distance to the squire center. This computation is described in equation 2.

$$d(x, y) = \sqrt[4]{x^4 + y^4} \quad (2)$$

As for PSCVC, a linear approach (Equation 1) uses this distance to compute the alpha factor. The distances that define minimal and maximal opacities are expressed in screen space.

As the alpha computation only depends on the 3D positions of the scene objects and to the corresponding projected pixel positions, it is totally independent of optional head tracking.

One possible result is given in Figure 1c where we apply our filter on the two spheres subject to frame cancellation. For this rendering we have set the minimal distance to 0.8 and the maximal one to 0.97. We can see on this figure that the virtual objects are smoothly clipped by the horizontal and the vertical screen edges; this clipping would have been rougher with state of the art approaches.

4 COLLISION DETECTION TO AVOID FRUSTUM CLIPPING

We have shown in Section 2 that there are many rendering adaptations to deal with frame cancellation, but only one method proposed by Wartell et al. [14] focuses on adapting the interaction to solve this issue. If we do not want to affect the scene rendering as in Section 3, this kind of solution is needed. In order to avoid frame cancellation, we propose to adapt the user interaction on the virtual objects by constraining the position of the manipulated object into the Stereo Compatible Volume (SCV) described by Ardouin et al. [1]. Indeed, when manipulating virtual objects, the user may place them in a conflict area corresponding to the outside of the red zones in Figure 2. That is why we propose a new approach to avoid such a situation during the object manipulation. Contrary to the method introduced by Wartell et al. [14], ours does not reduce the possibilities of creating the popping out effect during the manipulation.

4.1 Collision detection-based approach

Let us consider a user manipulating a virtual object in a stereoscopic co-located setup. In this case, the rendering depends on the user's inputs that define the motion of the manipulated object as well as the user's head position according to the display in order to ensure co-location. We attach to the manipulated object an encapsulating rigid body controlled by a physics engine detecting collisions with the two cameras frustums modeled with collision planes. The result is a collision detection between the object and the intersection of the two frustums.

We can split situations into 3 cases: (1) the user only moves a virtual object, (2) the user only moves his head, (3) the user moves both his head and a virtual object. In case (1), we simply apply a collision detection algorithm that ensures that the rigid body encapsulating the manipulated object never enters a conflicting zone. Case (2) is a little more tricky. In this case, we can apply two different policies. Either we let the physics engine directly handle the motion of the object if it is colliding a frustum plane, or we can control its motion, such as by fixing the object on the depth axis. This second solution constrains the physical behavior of the rigid body, avoiding the sliding of the object among the frustum plane assuming that the input stays fixed. In this case, if the user moves closer to the screen, the rigid body reacts as if it was pushed by the user's head. Actually, it is pushed by the intersection planes of

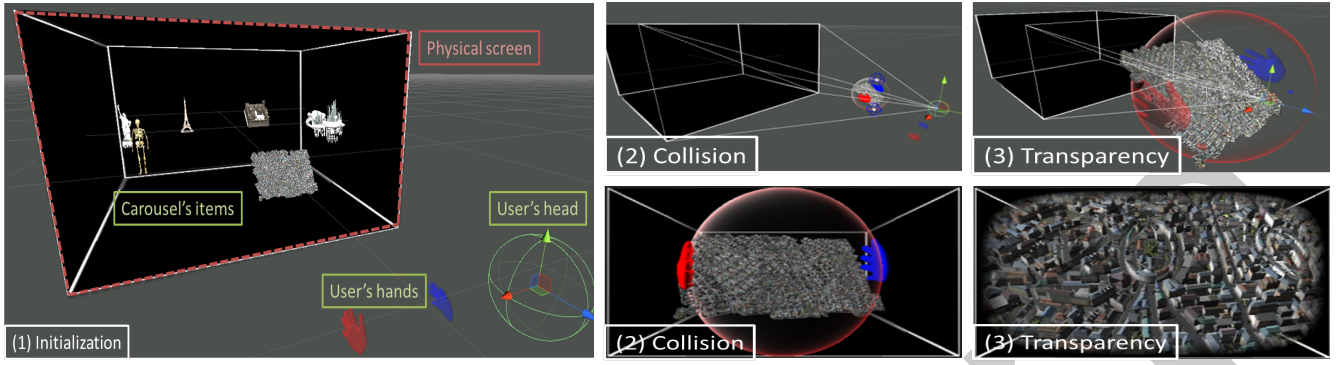


Figure 3: Illustrative use case (inspired from [6]) of the manipulation technique smartly merging the two proposed approaches to handle frame cancellation. The user is interacting with a virtual carousel carrying several virtual objects (1). The user can turn the whole carousel and manipulate the front object. The object can be a map that the user can bring very close to him with the collision detection-based approach activated (2). Then, he can switch the manipulation mode to scale the map and perceive small details with the virtual alpha blended window approach activated (3). The first line shows the scene from an external point of view while the second line is captured from the user's point of view.

the two frustums, which ensures that the object is always free of stereoscopic conflicts.

Case (3) is a merge of the two others, and actually is the only realistic one. Indeed, a user's input as well as a user's head position are continuously updated in real-time tracking systems. The adaptation of cases (1) and (2) to handle case (3) is trivial using an up-to-date physics engine. The only issue that can be overcome, is the management of collision detection and physical constraints by the physics engine when two dynamic user-controlled objects collide: the frustum planes defined by the user's head and the object defined by the user's input. This case is not trivially handled by physics engines, but can be managed thanks to additional controls in the physical loop. An implementation partially solving this issue consists of controlling the virtual object through a virtual spring attached to the user's input transform on one part and to the virtual object on the other part (inspired from [5]). This way, the virtual spring introduces an elastic aspect in the manipulation process. Well-customized, the virtual spring handles a smooth transition between the direct manipulation metaphor and the additional physical constraints managed by the physics engine.

This method dynamically adapts the interaction since we use physics to constrain some motion while manipulating an object. We can not control the range of the user's motion unless we use a haptic device with force feedback. Thus, our approach rather adapts the mapping of the user input regarding the object motion. In a non-conflicting zone, we can use a direct interaction metaphor (with or without an offset). If the system detects that the object is going to enter a conflicting zone, the interaction desired by the user is not ensured anymore since the motion is constrained to keep the object in a conflict-free area.

4.2 Limitations and extensions

If the size of a virtual object is taller than the screen's physical size, we can not move it in front of the screen since our algorithm is going to keep it behind the screen plane because of the collision with the frustums' planes.

A solution to overcome this limitation can be a smart merge of this collision detection-based approach with one of the effects presented in Section 3. For example, the user could enable a scaling mode in order to increase the size of an object and see it more detail. In that case, the collision detection algorithm may be disabled and one of the two rendering effects described in Section 3 may be enabled in order to reduce the frame cancellation effect, possibly caused by the scaling. An illustrative use case is provided in Figure 3.

5 EVALUATION

The evaluation had two main objectives: first, to compare and evaluate the two effects introduced in Section 3 with SCVC [1], which seems to be the best state of the art approach. All effects have been compared to a simple rendering in order to see if displaying a treatment is really necessary to deal with frame cancellation. Then, we evaluated the avoidance of frame cancellation with collision detection in a manipulation task as detailed in Section 4. To do so, we compared the technique with a non-constrained one in a basic scenario. The two experiments have been made by the same group of participants.

Participants Our experimental group consisted of 21 subjects aged from 20 to 48 (age: $M=34$, $SD=8$). There were 14 males and 7 females. They had little knowledge about rendering techniques and had very limited experience with stereoscopy. The subjects have various backgrounds: PhD students, R&D engineers, managers and assistants.

The Simulator Sickness Questionnaire We applied a Simulator Sickness Questionnaire (SSQ) at the beginning and end of the experiment [7]. The SSQ contains 16 physical symptoms rated on a categorical labeled scale (none, slight, moderate, severe). A factor analysis revealed that these symptoms can be placed into three general categories: oculomotor, disorientation, and nausea [7]. Weights are assigned to each of the categories and summed together to obtain a single score. The SSQ mean score was greater at the end than at the beginning of the experiment (start: $M=5.7$, $SD=1.2$; end: $M=6.4$, $SD=1.3$) mainly due to oculomotor symptoms. This increase was not significant. As the SSQ scores were low, we assumed that our VE did not produce strong cybersickness symptoms.

5.1 First step: Rendering techniques evaluation

5.1.1 Experimental apparatus

The application showed to the user different spheres of 15cm in diameter. The application background was blue and the spheres' color was red, as shown in Figure 1a. These colors were chosen in order to reduce the ghosting effect for the target display screen. Each sphere was different from the others only by the effect applied close to the screen edges. For the evaluation, four effects were compared, SCVC [1], progressive SCVC described in Section 3.1, the virtual alpha blended window detailed in Section 3.2 and a last effect, which applied no specific treatment. During the evaluation, these spheres moved on a plane that was one meter in front of the

screen and parallel to it. This distance was chosen in order to be included in the comfortable viewing zone for the current display setup with respect to the recommendations given by Chen et al. [4]. The sphere velocity (0.2 m/s) was chosen according to subjective pre-test results. The user was positioned seated with his head aligned with the screen center at about 2.80m distance to the screen. No head tracking was performed in order to modify the camera parameters; therefore, eyes were considered as X-aligned.

5.1.2 Hardware

We used a FaceLAB 5 eye-tracker system developed by Seeing Machines¹. This system allowed us to track head movement and eyes in real time. These data was not analyzed in real-time, but collected for analysis after the evaluation. It operated non-intrusively, with no head-mounted device. This was advantageous as the subjects had to wear polarized 3D glasses and sometimes their corrective glasses. The display screen was a Planar LCD, 2.08m diagonal, 4K with passive stereoscopy. Moreover, an Xbox wireless controller was used to interact with the application.

5.1.3 Procedure

The experiment started with a questionnaire and a Miles test [9] to evaluate dominant eye if unknown. In this test the observer creates a small opening with his hands. With both eyes, he looks at a distant object through this opening. The observer then alternates closing the eyes. The dominant eye is the one viewing the object. After this test, we calibrated the eye-tracker in precision mode for each subject.

Afterwards, for each side, the subjects saw each effect one after the other. Then, the subjects saw each effect for another side, and so on, one side after the other. They only had to visually follow the 3D sphere that moved about 1.80m in front of their face.

We applied a counterbalancing design with two between-groups variables: order of the side and order of the effects. *Order of the sides*: counterbalancing the order of the four sides leads to 24 orders. Each subject was assigned to a different order. For example [Down, Right, Down, Left] was assigned to subject 2. The participants did not know the order. *Order of the effects*: we called "series" an order of four effects. Counterbalancing the order of the four effects leads to 24 series. These series were assigned to the subjects. For example, the first series of subject 2 was [SCVC, NT, PSCVC, VABW], the fourth series of this participant was [VABW, PSCVC, NT, SCVC], and the first series of subject 16 was [NT, PSCVC, VABW, SCVC]. The participants did not know the series assigned.

After each series, each subject was requested to fill out a subjective questionnaire to evaluate the effects. To do so, the subject remained free to see the effects again, once again one by one. The subject used the Xbox wireless controller. Each of the four buttons A, B, X, and Y of the controller allowed the subject to switch between the four effects. A corresponded to No Treatment (NT), B to Stereo Compatible Volume Clipping (SCVC), X to Progressive SCVC (PSCVC), and Y to the Virtual Alpha Blended Window (VABW). The name-effect relations were identical during all of the experiment and for all subjects. The participants were just told about the effects' letters and did not know their names and specificities.

5.1.4 Data collection

In the subjective questionnaire, participants had to grade the four effects using a Likert-scale, from 1 (very low appreciation) to 7 (very high appreciation) according to five subjective criteria. They were defined by Arduin et al. ([1]): (a) Global appreciation, (b) Aesthetic, (c) Eye strain perceived, (d) Relief quality not at borders,

and (e) Relief quality at borders. The participants could choose NA (no answer). To facilitate the comparisons, we calculated the score of eyestrain such that the lower the eyestrain the higher the score. We called this new variable; "comfort according to eye strain perceived". To compare the five criteria, we calculated a weighted average for each subjective criterion. As some participants choose NA, we divided each weighted average by its theoretical maximum score (7 points multiplied by number of numeric answers).

Although the FaceLAB system tracked many head and gaze variables, for the purposes of this paper, we analyzed the gaze fixation point. The fixation point was computed automatically by FaceLAB as the intersection between the unified gaze ray and the work plane (the planar screen), defined relative to the screen coordinate system. This coordinate system was defined by the screen bottom left corner: point (0,0) and the top right one: point (1,1). We analyzed the relative distances ($D_x = G_x - O_x$, $D_y = G_y - O_y$) between the gaze fixation point (G_x , G_y) and the 3D object positions (O_x , O_y) in the same coordinates system. If the absolute value of this relative distance was up to a fixed limit (equal to 0.2), we considered that the subjects deflected attention away from the object. We also measured the ratio of deflected fixations which is the number of deflected fixations divided by the total number of fixations (deflected and non-deflected).

5.1.5 Hypothesis

Our hypothesis was that our two approaches described in Section 3, PSCVC and VABW, should give better results than the two other approaches, NT and SCVC effects:

- H1: Regarding Likert subjective questionnaire, we expect improved qualitative appreciations for the methods PSCVC and VABW. In particular, the global appreciation would be greater, and eyestrain would be lower.
- H2: The subjects could track the moving 3D objects close to the screen edges in a more comfortable manner with the methods PSCVC and VABW than the others. In others words, the subjects would be able to follow the 3D objects longer at the screen edges with the methods PSCVC and VABW than the other ones. Regarding specific variable, the ratio of deflected fixations just at the screen edges would be less for PSCVC and VABW effects than for NT and SCVC effects.

5.1.6 Results

Subjective criteria

Figure 4 shows the results concerning the grades (Likert-scale) obtained by the four different effects for each of the subjective criteria. The results for right and left sides were not significantly different as well as for top and bottom sides. Thus, only two graphics illustrate results.

The analysis of the subjects' responses gave two main results:

- For left and right sides: SCVC was the worst effect, VABW the best one,
- For up and bottom sides: NT, SCVC and PSCVC are equivalent and VABW was the worst effect.

For left and right sides, as we can see in Table 1, significant differences between the effects were found. SCVC was the worst effect for all the participants. SCVC was globally the least appreciated, the worst aesthetic effect, and presented the poorest relief quality at borders. A significant difference between the effects SCVC and VABW was found: SCVC provided stronger eyestrain than VABW. For top and bottom sides, few significant differences between the effects were found; in particular, VABW was globally less appreciated, as shown in Table 1. Regarding the order of preference, for each side, the participants were asked to order the effects from first to last place. Different effects could be equally placed.

¹ <http://www.seeingmachines.com/>

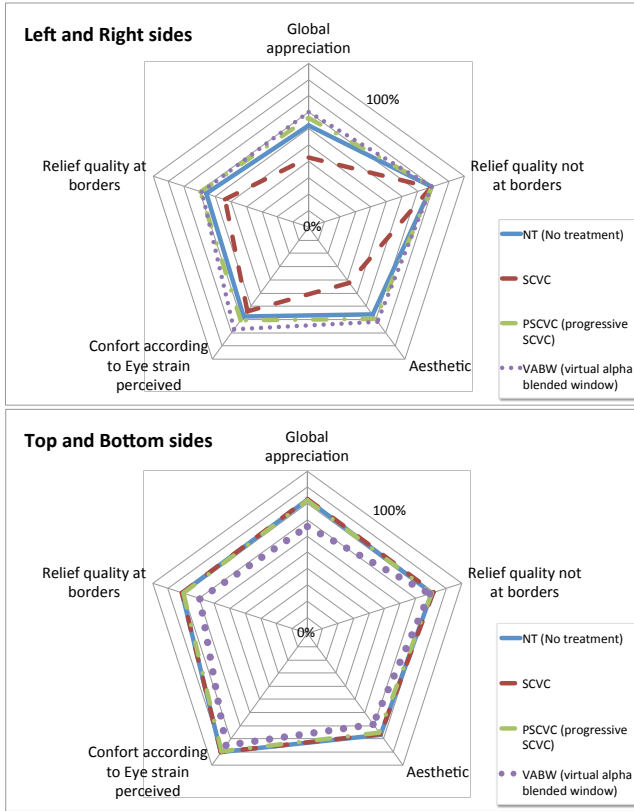


Figure 4: Results for the subjective questionnaire for the four different techniques (1) Control (blue), (2) SCVC (red), and our two new methods (3) PSCVC (grey), and (4) VABW (purple) with respect to Likert-scale grading

We analyzed the ranking order, in particular the frequency of effects chosen in the first two places. Significant difference between effects in the first two places was found ($\chi^2 = 18, df = 9, p < .05$). In comparison to the empirical probability distribution applied in the χ^2 test, SCVC was less preferred for right and left sides and PSCVC and VABW were more preferred for these two sides.

We calculated, for each side, the percent of times the participants ordered each effect in the first two places as shown in Figure 5.

For right and left sides, PSCVC and VABW were the most appreciated effects. For top and bottom sides, NT and SCVC were the most appreciated. We observed a gap between PSCVC, SCVC and NT for top and bottom but this gap was not significant. Indeed, the virtual object aspect was the same for the three effects on these sides.

Eye-Tracking results

To continue, concerning the eye-tracking results, we compared, for each side, the ratio of deflected fixations. We used R software² (lme4 library) and applied a mixed model to analyze this ratio just at the screen edges. Regarding our hypothesis, the ratio of deflected fixations just at the screen edges would be less for the PSCVC and VABW effects than for NT and SCVC effects.

For the right side, we obtained a significant difference between effects ($F(3,60) = 3.24, p < .05$). The ratio of deflected fixations was smaller with VABW than with the other effects. For left, top and bottom sides, we did not obtain any significant difference between methods.

For top and bottom sides, the result was not surprising. Indeed, as users' head rotation was not taken into account in the stereoscopic camera model, they did not perceive vertical parallax. Thus

χ^2 Test		Right and Left sides			Top and Bottom sides		
Df = 2; p value (χ^2 value)		SCVC	PSCVC	VABW	SCVC	PSCVC	VABW
NS= No Significant difference							
Global appreciation	NT	p<.001 (16,22)	NS	NS	NS	NS	p<.001 (21,52)
	SCVC		p<.001 (20,91)	p<.001 (25,54)		NS	p<.001 (23,81)
	PSCVC			NS			p<.001 (21,52)
Aesthetic	NT	p<.001 (21,2)	NS	NS	NS	NS	NS
	SCVC		p<.001 (25,55)	p<.001 (33,59)		NS	NS
	PSCVC			NS			NS
Eye strain perceived	NT	NS	NS	NS	NS	NS	NS
	SCVC		NS	p<.05 (8,5)		NS	NS
	PSCVC			NS			NS
Relief quality at borders	NT	p<.05 (7,6)	NS	NS	NS	NS	p<.01 (13,62)
	SCVC		p<.01 (9,66)	p<.01 (10,72)		NS	p<.01 (13,62)
	PSCVC			NS			p<.01 (11,19)
Relief quality Not at borders		All effects	NS			NS	

Table 1: Statistical results (χ^2 test) to compare the four methods

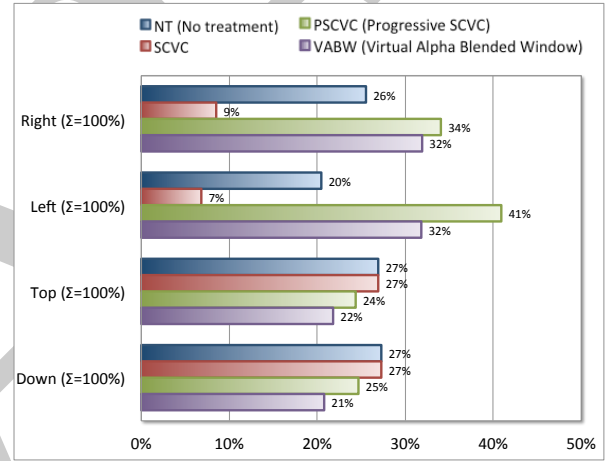


Figure 5: For each side, percent of times the participants ordered each effect in the first two places

NT, SCVC and PSCVC effects were visually similar for top and bottom sides. The result for left and right sides was more surprising. Considering the subjective appreciations, we hoped for a larger difference between the effects. To understand the reason for these results, we analyzed the effect of the dominant eye.

Significant difference between dominant eyes was found at the screen edges ($F(1,312) = 14.344, p < .001$). The ratio of deflected fixations was higher with left-eye dominant participants ($M=0.50, SD=0.08$) than with right-eye dominant ($M=0.25, SD=0.06$). The difference concerning each side is given in Figure 6. As a bit less than 2/3 of the participants (16) were right-eye dominant and 1/3 left-eye dominant (5), the effect of the dominant eye cannot be more discussed.

5.1.7 Discussion

Our hypotheses were not all verified.

H1: PSCVC and VABW improve qualitative appreciations
We found that participants clearly preferred PSCVC and VABW for right and left sides. In particular for global appreciation. For top and bottom sides, we found that the participants preferred NT and SCVC, but differences between the four effects were low.

As shown in Figure 4, PSCVC and VABW differ significantly from NT and SCVC for right and left sides. For top and bottom sides, in Figure 4 we also observe a little difference between VABW and the three other effects. These two observations can be explained

²<http://www.r-project.org/>

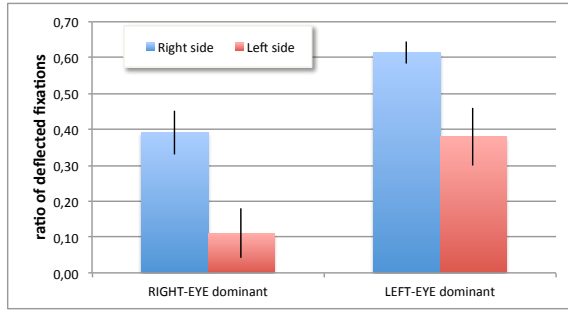


Figure 6: Ratio of deflected fixations

by binocular rivalry only present at the vertical edges in our experiment. Indeed, as our two effects were preferred on the vertical edges and no treatment was preferred on the horizontal ones, we can deduce that in order to deal with frame cancellation, binocular rivalry has to be especially considered. This consideration was done by the two effects we introduced and by the SCVC method. Regarding the results, the smooth aspect of our solutions seems to be preferred over the rough one proposed by SCVC. It is difficult to explain why NT was preferred over SCVC on the vertical edges because it contradicts the results obtained by Ardouin et al. [1]. We assume that the too rough clipping created a more disturbing effect than binocular rivalry and frame cancellation. However, this result could also be explained by the difference between our experimental setup and the one used by Ardouin et al. [1]. More work could be done in that way to compare the impact of different display setups and different VEs.

H2: PSCVC and VABW improve 3D object tracking at the screen edges The number of deflected fixations just at the screen edges for right side was smaller for VABW. Thus, the participants could track the moving 3D objects close the screen's right edge in a more comfortable manner. We found that VABW was better than the other effects for right side. But, we did not obtain other significant results without taking into account the dominant eye.

Indeed, the dominant eye played an important role in the deflecting movement. Left-eye dominant participants showed more deflected fixations than the others. We did not control eye dominance as an independent variable in the experiment, but our initial results suggest effects regarding the physiological reactions to 3D moving objects at the screen edges, which warrant future studies in this direction.

5.2 Second step: Interaction technique evaluation

5.2.1 Experimental apparatus

In the second experiment, the application displayed four artworks, four pedestals and a wall as shown in Figure 7. The goal was to put each artwork on its corresponding pedestal with a 3D-ray based manipulation technique. This technique is implemented as described by Bowman and Hodges [3] where a light ray extends from the user's hand detected by a 6-DoF tracker. When the ray intersects a scene object and the user presses a selection button, this object is attached to the ray extremity. The pedestals were aligned horizontally. Each artwork was identified by a number at its base. To find the corresponding pedestal of each one, the user had to sort them in ascending order. The lowest number corresponded to the leftmost pedestal, the highest to the rightmost one. In order to make sure that situations with frame cancellation occurred, we designed the experiment such that users had to bring the objects close to themselves during the trials. First, the wall was placed between the pedestal and the initial position of the artwork. It was a physical barrier that forced the user to move the objects in front of it to put them in the pedestals (the height of the wall was virtually infinite but not entirely represented in order to avoid an initial frame cancellation



Figure 7: The four masterpieces separated from the pedestals by a wall

situation). Its length allowed us to control the minimal popping-out distance. Second, the numbers written on the bases of the artworks had two digits and were so small that they needed to be really close to the user to be well read.

5.2.2 Hardware

For this experiment we used the zSpace interactive system which is composed of a passive stereo display (24-inches) with head tracking for co-location, and a stylus which provided a 6-DoF tracking and buttons that were used for object grabbing. Moreover, we added a numeric keypad to this system. At any time, a subject could use a key of this device to reset the position of the virtual objects, for example, when an object got lost. We did not use a stylus key for this feature in order to avoid accidental actions for novice users.

5.2.3 Procedure

The experiment started with a training phase. For five minutes, the user trained himself how to move the virtual objects with the 3D-ray and how to place them on the different pedestals. Once this step was over, the participant performed two trials to compare our approach called Conflict-Free (CF) mode with the non-constrained one called Liberty (L) mode. We applied a counterbalanced design: $2 \text{ modes} \times 2 \text{ series}$. **Modes:** Half of the participants began with CF, half with L. The users were not told about the names and the characteristics of the two modes. **Series:** we defined two series of numbers on the artworks. Each subject was assigned to one of the four conditions. For example, the first subject followed L mode with series 1 then CF mode with series 2.

After the two trials, each subject was requested to express their subjective impression in terms of usability and visual comfort.

5.2.4 Data collection

We collected the following data: number of manipulations, time to success, time and number of manipulations to put the first object in the pedestals, time and number of manipulations to put the fourth object in the pedestals (the first try).

5.2.5 Hypothesis

Since CF constraints the object motion to keep it in a conflict-free area, we made two assumptions.

- H1: Conversely to L mode, the participants in CF mode could not up-scale an object too much in front of them and could not generate frame-cancellation situations. We hoped that users would express more visual comfort with CF mode.
- H2: The participants in CF mode could not lose a part of an object culled by the cameras frustums so the performance should be better.

5.2.6 Results

We analyzed all the data. Only one significant difference between the two modes was found: the number of manipulations to put the

fourth object in the pedestals ($F(1,40) = 42, p < 0.05$). The participants in CF made fewer manipulations ($M=10.3$, $SD=2.1$) than in L mode ($M=12.1$, $SD=3.9$). Thus, participants could complete their task in the same manner in the two modes. The L mode did not constrain the interaction of the participants. In contrast, the CF mode optimized the number of manipulations until the first try.

Even if the difference of number of manipulations to put the fourth object in the pedestals was significant, the participants did not express it often in the questionnaire at the end of the experiment. 2/3 of the participants (14) did not notice any functional differences between the two modes. In the 1/3 of the participants (7), a few did not appreciate the co-located manipulation technique without constraint because the objects did pop out too quickly: “*The object becomes too big too fast*”. In contrast, a few participants did not appreciate the CF mode for the inverse reason; they were not able to pop out the virtual objects enough: “*It is less easy to pull the object in this mode*”.

5.2.7 Discussion

The performances have improved a little with this mode (H2 verified). We assume that the CF mode helped and guided the users to complete the manipulation task. Future work may focus on comparing the two modes in another scenario in order to confirm the results on the interaction performances. Regarding H1, the participants did not express more visual comfort in CF mode in the qualitative results (H1 not verified). These results might be explained by the short duration (5-10 minutes) of the experiment. We assume that we would have found clearer results with a longer duration. More work should be done to evaluate the impact of the duration on the visual comfort with the two modes.

6 CONCLUSION

In this paper, we introduced different solutions to deal with frame cancellation in 3D user interfaces.

In Section 3, we introduce two rendering techniques that try to solve the main drawbacks of current solutions. Indeed, the state of the art approaches suffer from limitations such as field of view reduction, scene scale alteration or too rough clipping of virtual objects. The two solutions proposed are based on a progressive alpha blending to let the objects disappear close to the physical screen edges. The first technique, progressive stereo compatible volume clipping (PSCVC), is mainly based on stereo compatible volume clipping (SCVC) [1] and proposes to apply it progressively with an alpha blending in order to not disturb the observer with a rough clipping. The second one, the virtual alpha blended window (VABW) computes its alpha factor for the two eyes based on the distance of each pixel to the closest screen border. Therefore, this technique takes into account vertical borders as well as horizontal ones. Our evaluation shows that according to subjective criteria VABW is the best technique for left and right sides but is the worst for the top and bottom ones. PSCVC presents good results independently to the target screen side. We observed a bigger impact of frame cancellation on the horizontal edges, which can be explained by the additional problem of binocular rivalry present on these sides. We also observed surprisingly low results for SCVC, which is the least appreciated effect on the left and right sides. Future studies are needed in order to understand this result in the scope of different display setups (passive/active, display size) and different VEs. Our eye tracking data suggests that participants tracked the moving 3D objects close to the right screen edge in a more comfortable manner with the VABW effect. We did not obtain any significant results for the other edges. Moreover, we observed a potential impact of eye dominance on the results. However, due to the small number of left-eye dominant participants in our experiment, future studies are required to evaluate this aspect in detail. Future work could also

compare the different effects with more complex objects in more complex scenes.

Then, we also introduce a new approach designed for 3D user interactions. Instead of adapting the rendering of objects in frame cancellation areas, the technique proposes to constrain the interaction in order to avoid this kind of situation. The technique is based on a collision detection system in order to constrain the manipulated object in the stereo compatible volume described by [1]. The evaluation shows that the users did not notice visual comfort differences between this mode and a non-constrained one, but their performances improved a little with the constrained mode. We surmise that with longer experiments, the user may notice a difference between the two modes in terms of interaction quality and visual comfort. Future work may focus on evaluating this method in more realistic scenarios in order to confirm its efficiency.

REFERENCES

- [1] J. Ardouin, A. Lécuyer, M. Marchal, and E. Marchand. Design and Evaluation of Methods to Prevent Frame Cancellation in Real-Time Stereoscopic Rendering. In *IEEE Symposium on 3D User Interfaces, 3DUI 2011*, pages 95–98, Singapore, Singapore, 2011.
- [2] Autodesk. *Stereoscopic filmmaking whitepaper, The Business and Technology of Stereoscopic Filmmaking*. 2008.
- [3] D. A. Bowman and L. F. Hodges. An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments. In *Proceedings of the 1997 Symposium on Interactive 3D Graphics, I3D '97*, pages 35–ff., New York, NY, USA, 1997. ACM.
- [4] W. Chen, J. Fournier, M. Barkowsky, and P. Le Callet. New stereoscopic video shooting rule based on stereoscopic distortion parameters and comfortable viewing zone. In *Stereoscopic Displays and Applications XXII, SPIE 2011*, pages SPIE 7863, 78631O, San Francisco, United States, Jan. 2011.
- [5] B. Frohlich, H. Tramberend, A. Beers, M. Agrawala, and D. Baraff. Physically-based manipulation on the responsive workbench. In *Virtual Reality, 2000. Proceedings. IEEE*, pages 5–11. IEEE, 2000.
- [6] P. Gaucher, F. Argelaguet, J. Royan, and A. Lécuyer. A novel 3d carousel based on pseudo-haptic feedback and gestural interaction for virtual showcasing. In *3D User Interfaces (3DUI), 2013 IEEE Symposium on*, pages 55–58. IEEE, 2013.
- [7] R. S. Kennedy, N. E. Lane, K. S. Berbaum, and M. G. Lilienthal. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology*, 3(3):203–220, 1993.
- [8] L. Lipton and B. Akka. Vertical surround parallax correction, Mar. 16 2010. US Patent 7,679,641.
- [9] W. R. Miles. Ocular dominance in human adults. *The journal of general psychology*, 3(3):412–430, 1930.
- [10] J. D. Mulder and R. V. Liere. Enhancing fish tank vr. In *Proceedings of Virtual Reality*, pages 91–98. IEEE, 2000.
- [11] M. Slater. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535):3549–3557, 2009.
- [12] N. Valys. *Stereoscopy*. Focal library. Focal Press, 1966.
- [13] C. Ware and D. Fleet. Integrating flying and fish tank metaphors with cyclopean scale. In *Computer Graphics International, 1997. Proceedings*, pages 39–46. IEEE, 1997.
- [14] Z. Wartell, W. Ribarsky, and L. Hodges. Third-person navigation of whole-planet terrain in a head-tracked stereoscopic environment. In *Virtual Reality, 1999. Proceedings., IEEE*, pages 141–148, Mar 1999.
- [15] Z. J. Wartell. Stereoscopic head-tracked displays: analysis and development of display algorithms. 2002.